

Herl, H. August 1990. User's Manual, HyperCard Database for the Natural Language Sourcebook. Center for Technology Assessment, UCLA Center for the Study of Evaluation.

Read, W., Dyer, M., Baker, E., Mutch, P., Butler, F., Quilici, A., & Reeves, J. 1990. Natural Language Sourcebook. Center for Technology Assessment, Center for the Study of Evaluation and Computer Science, Department, UCLA.

**DESIGNING A HYPERCARD DATABASE
FOR THE NATURAL LANGUAGE SOURCEBOOK**

Howard Herl

**Center for Technology Assessment
UCLA Center for the Study of Evaluation**

September 1990



**Artificial Intelligence Measurement System
Contract Number N00014-86-K-0395**

Principal Investigator: Eva L. Baker

**Center for Technology Assessment
UCLA Center for the Study of Evaluation**

This research report was supported by contract number N00014-86-K-0395 from the Defense Advanced Research Projects Agency (DARPA), administered by the Office of Naval Research (ONR), to the UCLA Center for the Study of Evaluation. However, the opinions expressed do not necessarily reflect the positions of DARPA or ONR, and no official endorsement by either organization should be inferred. Reproduction in whole or part is permitted for any purpose of the United States Government.

Approved for public release; distribution unlimited.

Accession For	
NTIS	<input checked="" type="checkbox"/>
DTIC	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and/or Special
A-1	

The author wishes to thank Frances Butler, Harold F. O'Neil, Jr. and Patricia Mutch for their comments and suggestions during the preparation of this report.

Contents

Initial Section.....	v
Background.....	v
Description of the Final Report.....	v
Project Deliverables.....	v
Purpose of the Sourcebook Database.....	v
About the Natural Language Sourcebook	vi
Section 1 Specification of System Components.....	1
System Components.....	1
Introduction to the Sourcebook.....	1
Exemplars.....	1
Sourcebook Classification Scheme.....	3
Linguistic Classification Scheme.....	14
Cognitive-Psychological Classification Scheme.....	15
Complete References.....	16
Appendix.....	16
Section 2 Design and Development of the Sourcebook Database.....	17
Sourcebook Database Specifications.....	17
Sourcebook Database Design.....	17
Integrating System Components.....	18
Browsing and Searching Features.....	11
On-Line Assistance.....	14
Sourcebook Database User's Manual.....	15
Summary.....	15
References.....	17
	111

Contents

Final Section

Background

The UCLA Center for Technology Assessment of the Center of the Study of Evaluation has an existing contract with the Defense Advance Research Projects Agency to study human benchmarking methodology for the evaluation of artificial intelligence systems. Systems of interest include vision, natural language, experts systems shells and expert systems. This final report documents the implementation of the Natural Language Sourcebook hardcopy document to a user friendly, interactive HyperCard database stack.

Description of the Final Report

The two primary tasks of the project were: 1-Specifying system components and 2-Designing and developing the Sourcebook database. The tasks of the project are described in the following sections:

- Section 1, "Specification of System Components," provides a description of the content in the Natural Language Sourcebook which is included in the electronic version of the Sourcebook database.
- Section 2, "Design and Development of the Sourcebook Database," describes the overall design and documentation of the development of the HyperCard stack for the Sourcebook database and the accompanying User's Manual.

Project Deliverables

In the Natural Language Sourcebook HyperCard Project there were three deliverables:

- Sourcebook Database
- Sourcebook Database User's Manual
- Final Report

Portions of the text and figures from the User's Manual have been included in this final report to provide precise explanations.

Purpose of the Sourcebook Database

The Sourcebook Database was developed to assist artificial intelligence researchers, cognitive scientists and computational linguists in accessing the vast amount of information contained in the Sourcebook in a more efficient and effective manner than possible using the hardcopy version.

About the Natural Language Sourcebook

The Natural Language Sourcebook is a collection of examples representing a range of processing problems in natural language understanding. The focus of the Sourcebook is on processing problems involved in building artificial intelligence systems. The processing problems emerged from a literature search primarily in the field of artificial intelligence but also in the related fields of computational linguistics and cognitive science.

Each entry in the Sourcebook represents a problem. The entries, called "exemplars," each consists of 1) one or more sentences, a fragment of dialogue, or a piece of text which illustrates a conceptual issue, 2) a reference, and 3) a discussion of the problem a system might have in understanding the example. An example is used to illustrate each problem, but it is the discussion that defines the type of problem by delineating the information-processing issues involved.

The Sourcebook is organized around a classification scheme which reflects an artificial intelligence perspective of the problems encountered in processing natural language. Additional classification schemes, one with a linguistic orientation and one with a cognitive-psychological orientation, provide the basis for cross-referencing the processing problems. Linguistic and cognitive-psychological cross-referencing of the exemplars allows the user to explore alternate relations among the examples represented in the Sourcebook. Each exemplar has at least one linguistic cross-reference and one cognitive-psychological cross-reference.

The Natural Language Sourcebook can serve as part of an evaluation methodology for artificial intelligence systems in that it provides an overview of natural language processing problems against which such systems can be assessed. The processing problems delineated by the Sourcebook classification scheme define four discourse types which an artificial intelligence system should ideally be able to handle. Those discourse types, the four major categories in the Sourcebook classification scheme, are: I. Single-utterance issues, II. Connected-utterance issues, III. True-dialogue issues, and IV. Ill-formed input.

Group I includes input which the system can understand using only existing, stored knowledge. This knowledge might be knowledge about syntax, semantic constraints, or world knowledge, but it specifically excludes knowing the results of processing previous input. The typical Group I problem involves parsing an input in isolation from other inputs. Most of the examples in this group are single sentences, but it should be noted that a "single-utterance" may consist of more than one sentence. The examples in Group I are the ones most commonly discussed in the linguistics and artificial intelligence literature.

In Group II problems, the system must be able to integrate information spread over a series of utterances and must be able to refer back to earlier utterances, but the system need not have any model of the user. The only interaction between the system and the user is that the user makes statements or requests, and the system processes them.

In Group III problems, the system must know not only about the sequence of utterances read but also something about the user's goals, intents, and expectations. Group III examples do not necessarily involve an actual two-way dialogue between the system and the user, but they

at least involve building or accessing a model of the user. Group III problems involve an interchange between two knowledgeable entities. Group III systems are needed for intelligent interfaces to expert systems.

Groups I, II, and III form an important block because they cover examples of normal language use. The remaining group, Group IV, ill-formed input, covers examples which do not conform to normal language use but which a program may nevertheless encounter.

(The discussion above was taken from Chapter 1, Introduction to the Sourcebook, in the User's Manual.)

Initial section

The first task for the implementation of the Natural Language Sourcebook into the HyperCard environment was to specify the essential content features of the Natural Language Sourcebook. These would become the system components for the database in the HyperCard environment. Seven components were identified: Sourcebook Introduction, Exemplars, Sourcebook Classification Scheme, Linguistic Classification Scheme, Cognitive-Psychological Classification Scheme, Complete References and Appendix.

System Components

- **Introduction to the Sourcebook**

The introduction to the Natural Language Sourcebook is presented in the initial section of this report.

- **Exemplars (197)**

Each of the 197 exemplars in the Natural Language Sourcebook contains the following six components (see Figure 1.1):

- **Exemplar Classification Type**
The exemplars are categorized into four major discourse types--I. Single utterance issues, II. Connected-utterance issues, III. True-dialogue issues, and IV. Ill-formed input--and various subtypes in the Sourcebook Classification Scheme.
- **Exemplar Identification Number**
The exemplar identification number corresponds to the placement of the exemplar in the Sourcebook Classification Scheme.
- **Problem Section**
This section consists of one or more sentences, a fragment of dialogue, or a piece of text which illustrates a conceptual issue.
- **Abbreviated Reference**
The abbreviated reference displays the author's name, date and page numbers of the reference from which the examples in the problem section were taken.
- **Discussion Section**
This section provides a discussion of the problem a system might have in understanding the example given in the Problem Section.
- **Linguistic and Cognitive-Psychological Cross-References**
These cross-references for each exemplar allow the user to explore alternate relations among the examples represented in the Natural Language Sourcebook.

■ FIGURE 1.1

Typical Exemplar

Exemplar classification

Exemplar identification number

Problem section

Abbreviated reference

Discussion section

Linguistic
classification entries

Cognitive-psychological
classification entries

SINGLE-UTTERANCE ISSUES

Identification of Syntactic Units

Exemplar I.A. (#2)

Problem

- (1)...mental building block
- (2)...large Chinese restaurant
- (3)...meat shop owner
- (4)...fearless Chinese soldier
- (5)...heavy cigar smoker
- (6)...Indian Ocean cruise line
- (7)...giant shrimp cocktails

(Gershman, 1982, pp. 182-185)

Discussion

In the noun groups above, there are some complications in attaching noun-modifiers to nouns. In "heavy cigarsmoker," "heavy" modifies "cigarsmoker" not "cigar." In "Indian Ocean cruise line," however, "Indian" modifies "Ocean" and not "Ocean cruise line."

The following principles help the parser determine the attachment of modifiers to nouns:

- 1) Only concepts built by adjacent words can modify each other: "mental block" is not the same as "mental building block."

Linguistic:

- I.A.1. Noun/verb
- I.A.2. Noun/adjective
- I.A.3. Noun/participial adjective
- I.C.1. Noun phrase structure

Cognitive-psychological: I.A. Processing syntax

• Sourcebook Classification Scheme

Figure 1.2 displays the Sourcebook classification scheme around which the Natural Language Sourcebook is organized.

■ **FIGURE 1.2**
Sourcebook
Classification Scheme

Sourcebook
classification
entries

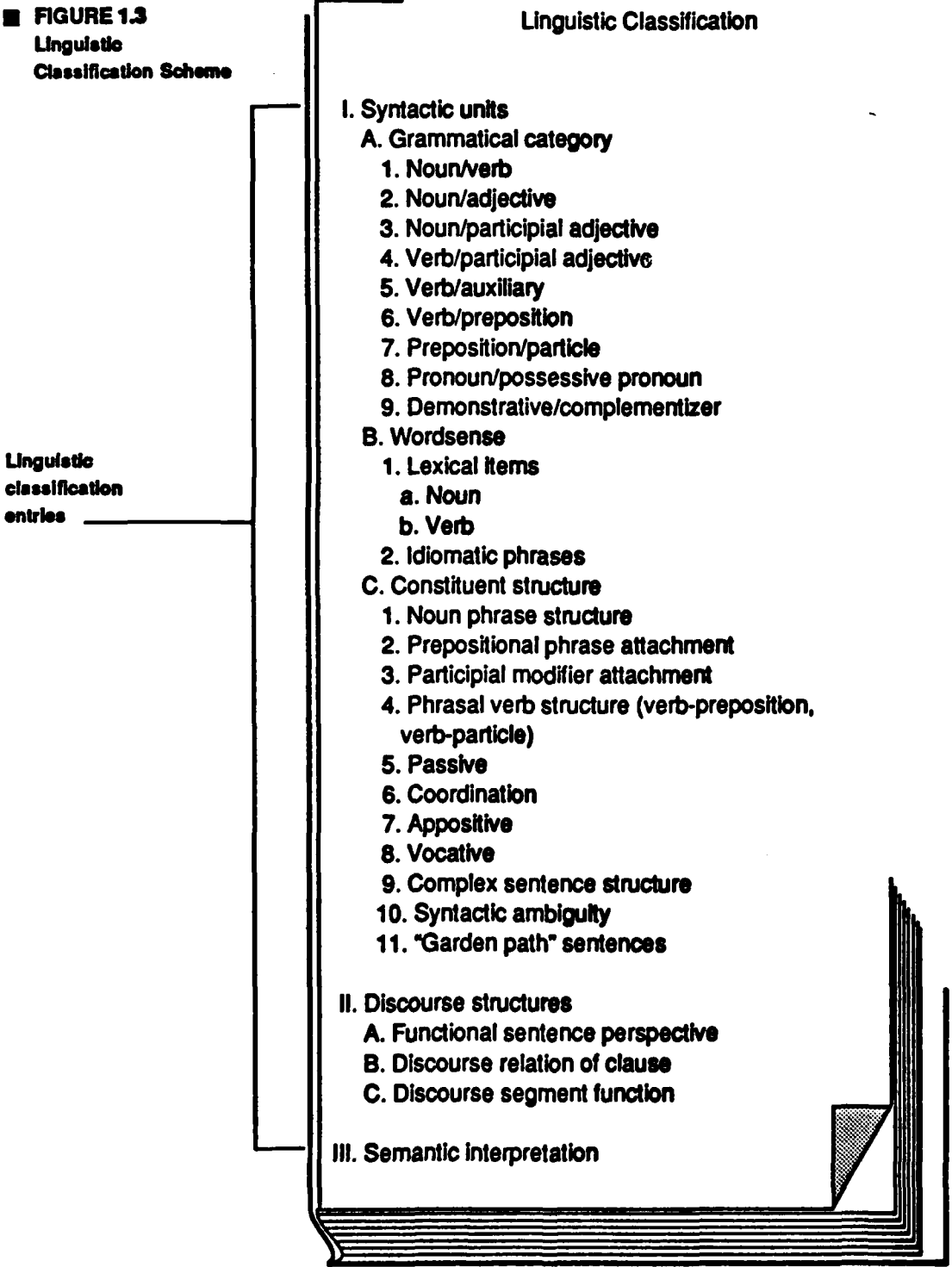
Sourcebook Classification

- I. Single-utterance issues**
 - A. Identification of syntactic units**
 - B. Ambiguity**
 1. Lexical
 2. Attachment
 - C. Modifier attachment**
 1. Prepositional phrase
 2. Other
 - D. Reference**
 1. Anaphoric
 2. Non-anaphoric
 3. Distinguishing anaphoric from non-anaphoric reference
 4. Temporal
 - E. Metaphor and novel language**
 - F. Other**
- II. Connected-utterance issues**
 - A. Anaphora**
 - B. Ellipsis**
 - C. Integrating complex information**
 - D. Reasoning, argumentation, story understanding**
 1. Irony
 2. Plans and goals
 3. Belief modification
 4. Argumentation
 5. Summarizing
 - E. Metaphor**
- III. True-dialogue issues**
 - A. User goals and plans**
 - B. Logical presuppositions**
 - C. Speech acts**
 - D. Meta-linguistic discourse**
- IV. Ill-formed input**
 - A. Mistakes**
 1. Mistypings
 2. Syntactic constraint violations
 3. Semantic constraint violations
 - B. Non-standard input**
 1. Incomplete sentence
 2. Casual structure
 3. Idiolect

• **Linguistic Classification Scheme**

Figure 1.3 displays the first of two pages used for the linguistic classification scheme.

FIGURE 1.3
Linguistic
Classification Scheme



• **Cognitive-Psychological Classification Scheme**

Figure 1.4 displays the first of 11 pages used for the cognitive-psychological classification scheme.

■ **FIGURE 1.4**
Cognitive-Psychological
Classification Scheme

Cognitive-
psychological
classification
entries

Cognitive-psychological Classification

- I. Processing single utterances
 - A. Processing syntax
 - 1. Type of sentences
 - a. Declarative
 - 1.) Active
 - 2.) Passive
 - b. Interrogative
 - c. Imperative
 - 2. Types of phrases and clauses
 - a. Nouns vs. modifiers
 - b. Noun groups
 - c. Prepositions in noun groups
 - d. Appositives
 - e. Other modifiers
 - 3. Word order
 - a. Standard
 - b. Inverted
 - 4. Referents
 - a. Structure-based
 - 1.) Of adjectives
 - a.) Post nominal
 - (1.) Participial phrases
 - (2.) Relative clauses
 - b.) Standard
 - 2.) Of pronouns
 - a.) Direct
 - b.) Indirect
 - 3.) Of objects
 - 4.) Of phrases
 - b. Frame or script-based
 - 1.) Definite references
 - a.) Referential
 - b.) Attributive
 - c.) Set
 - (1.) Generic
 - (2.) Individual

- Complete References of citations in the exemplars

Figure 1.5 displays part of the first of 9 pages used for the complete references of authors cited in the Natural Language Sourcebook.

■ FIGURE 1.5
References

Complete
reference
entries

References

Allen, J. F. (1987). Natural Language Understanding. Menlo Park, CA: Benjamin/Cummings.

Allen, J. F., Frisch, A. M., & Litman, D. J. (1982). ARGOT: The Rochester dialogue system. Proceedings of the National Conference on Artificial Intelligence, 66-70.

Allen, J., & Perrault, C. R. (1980). Analyzing intention in utterances. Artificial Intelligence, 15, 143-178.

Alvarado, S., Dyer, M. G., & Flowers, M. (1985). Memory representation and retrieval for editorial comprehension. Proceedings of the Seventh Annual Conference of the Cognitive Science Society, 228-235.

Appelt, D. E. (1985). Planning English referring expressions. Artificial Intelligence, 26, 1-33.

Arens, Y. (1986). CLUSTER: An approach to modeling context. Proceedings of the Eighth Annual Conference of the Cognitive Science Society, 595-600.

August, S., & Dyer, M. G. (1985). Understanding

- Appendix

The appendix provides discussions about the development of the linguistic and cognitive-psychological classification schemes.

Design and Development of the Sourcebook Database

The purpose of the second task was to provide the overall design and documentation of the development of the HyperCard stack for the Sourcebook Database and the accompanying User's Manual.

Sourcebook Database Specifications

The Sourcebook Database is a HyperCard stack which contains the information found in the Natural Language Sourcebook. The Sourcebook Database stack works with any Apple Macintosh computer with HyperCard 1.2 or higher, one megabyte of memory or more and at least one 800K disk drive. The Sourcebook Database stack contains 335 cards, which have been classified into the following categories:

<u>Description</u>	<u>no. of cards</u>
Title	1
Introductory	4
Exemplars	197
Classification Schemes	40
Complete References	91
Quick Reference	1
Exit	1

Sourcebook Database Design

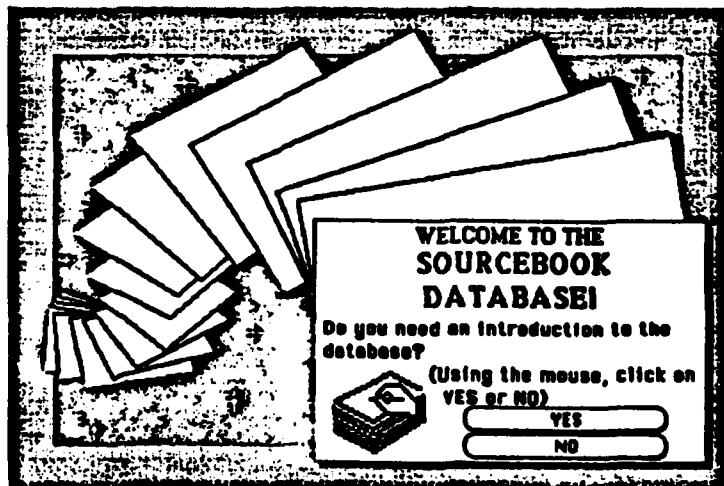
The Sourcebook Database was designed into three modules: 1) Integrating system components, 2) Complementing the system components with browsing and searching features, and 3) Incorporating on-line assistance features to accomplish user-friendliness.

- **Title card**

The opening card welcomes the user to the Sourcebook Database and asks the user if an introduction to the database is requested (see Figure 2.1)

■ **FIGURE 2.1**

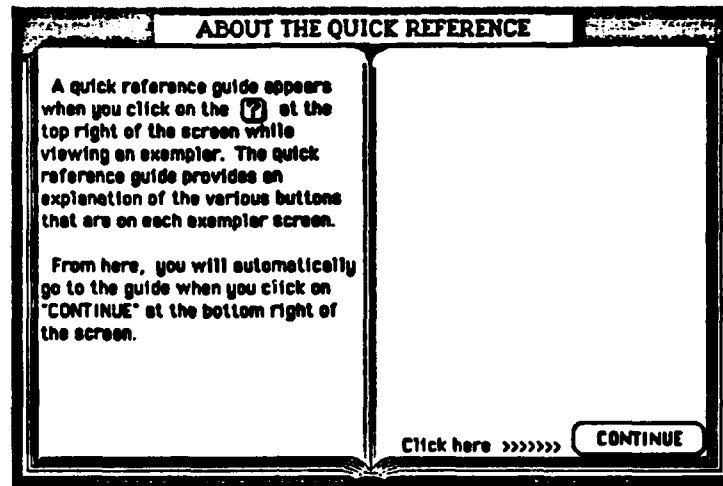
Title Screen



- **Introductory cards**

The introductory cards contain information about the project, Natural Language Sourcebook, Prompting mode and Quick Reference Guide (see Figure 2.2).

■ **FIGURE 2.2**
About the Quick
Reference Guide

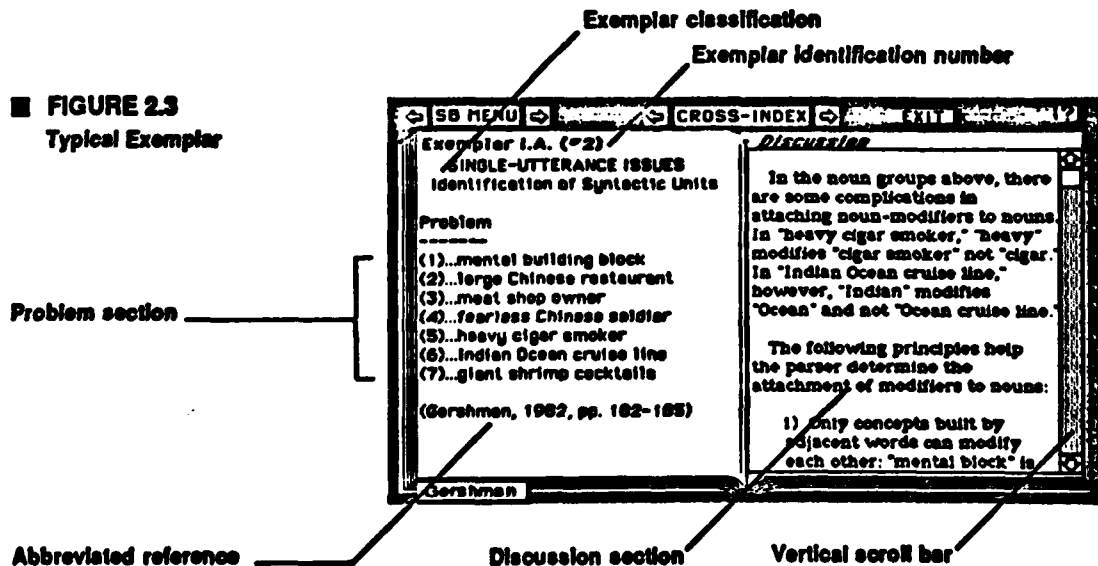


Integrating System Components

- **Exemplars**

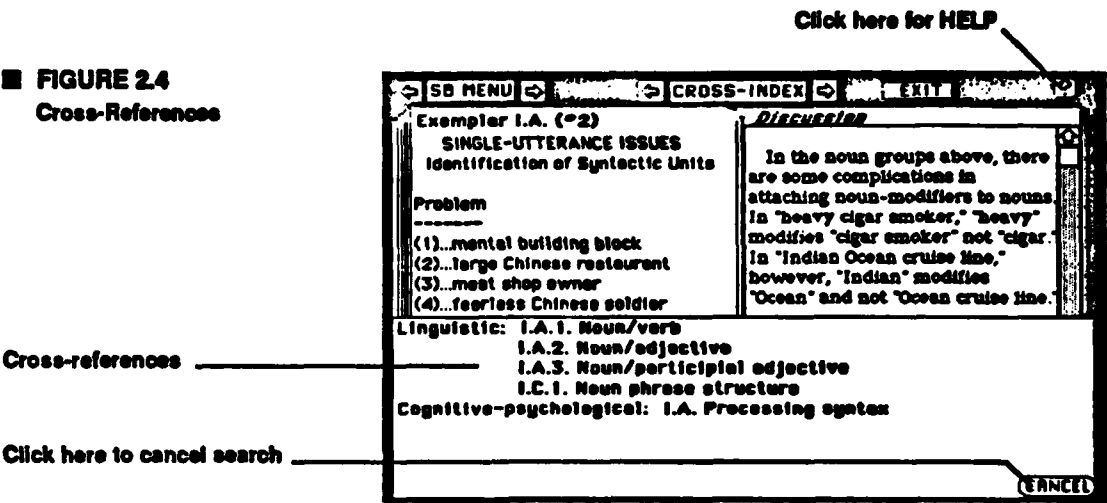
Each of the 197 exemplars is displayed on one card in the format of an open book (see Figure 2.3). The left page of each exemplar contains the exemplar identification number, type of exemplar classification, problem section section, and abbreviated reference. The right page contains the discussion section only. The problem and discussion sections are presented in scrolling fields when they are longer than one screen.

■ **FIGURE 2.3**
Typical Exemplar



Each of the exemplars has at least one linguistic and one cognitive-psychological cross-reference. These cross-references are displayed when the user clicks on either of the cross-index arrows (see Figure 2.4).

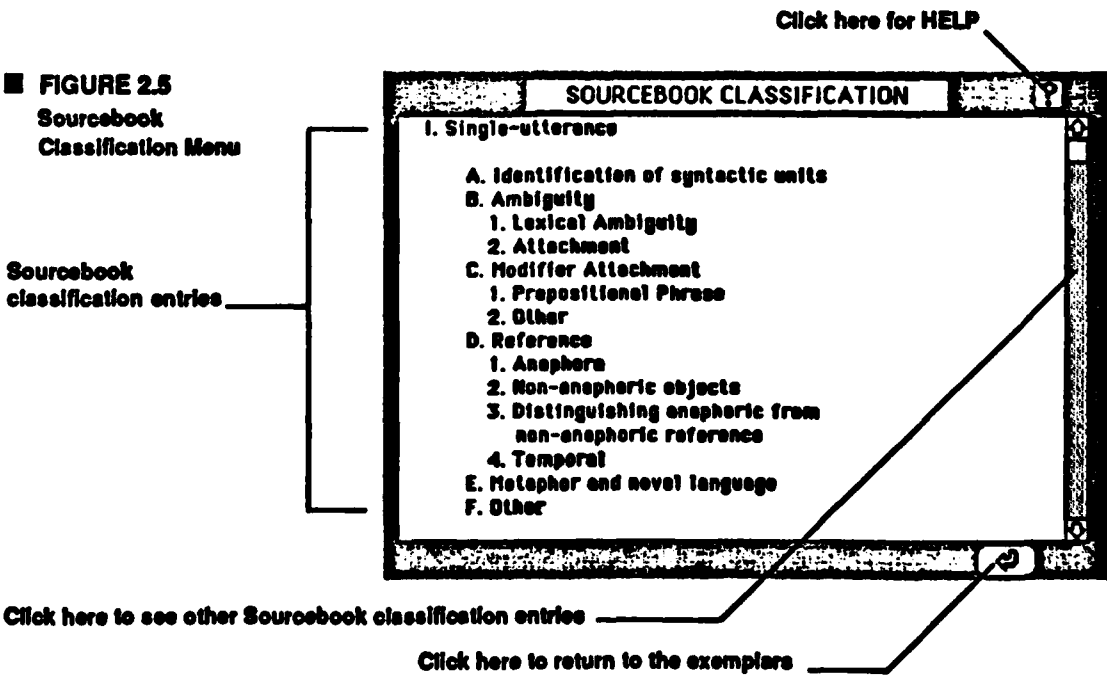
■ FIGURE 2.4
Cross-References



• Sourcebook Classification Scheme

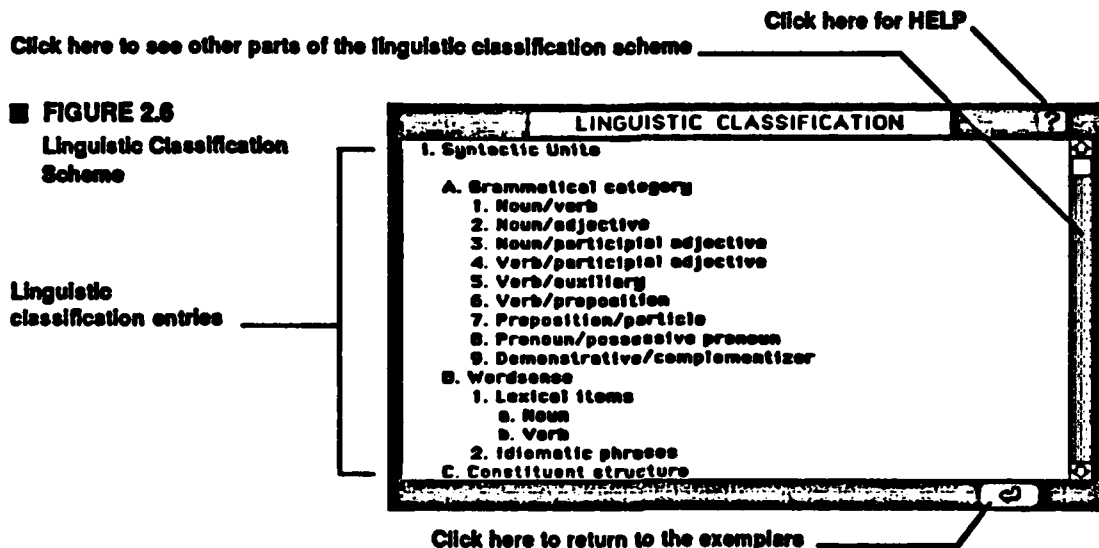
The Sourcebook classification scheme is displayed on one card and is contained in a scrolling field (see Figure 2.5).

■ FIGURE 2.5
Sourcebook
Classification Menu



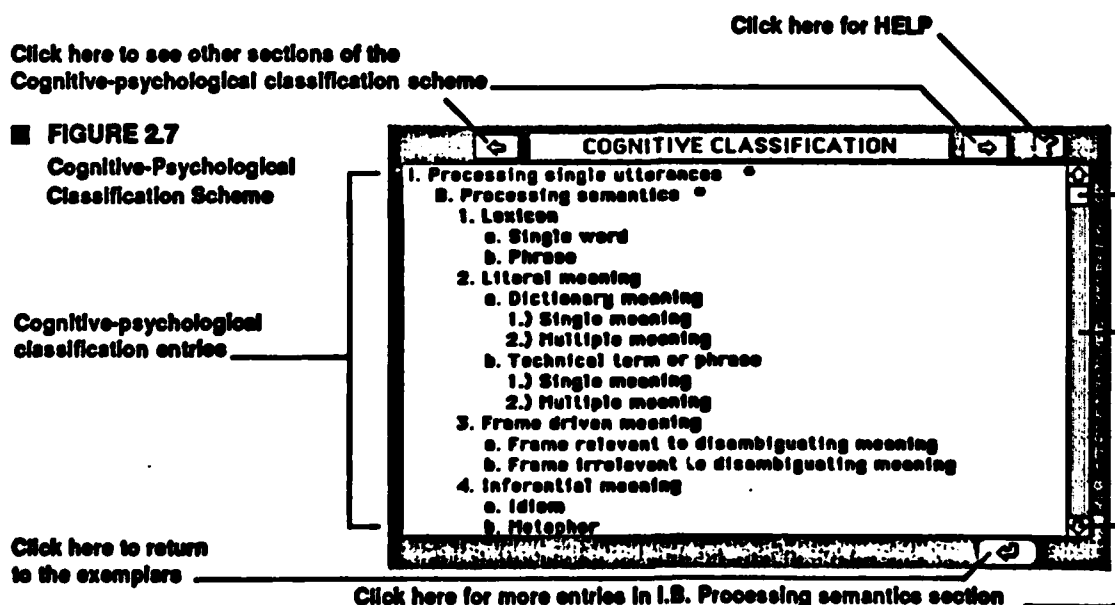
• Linguistic Classification Scheme

The linguistic classification scheme is displayed on one card and is contained in a scrolling field (see Figure 2.6).



• Cognitive-Psychological Classification Scheme

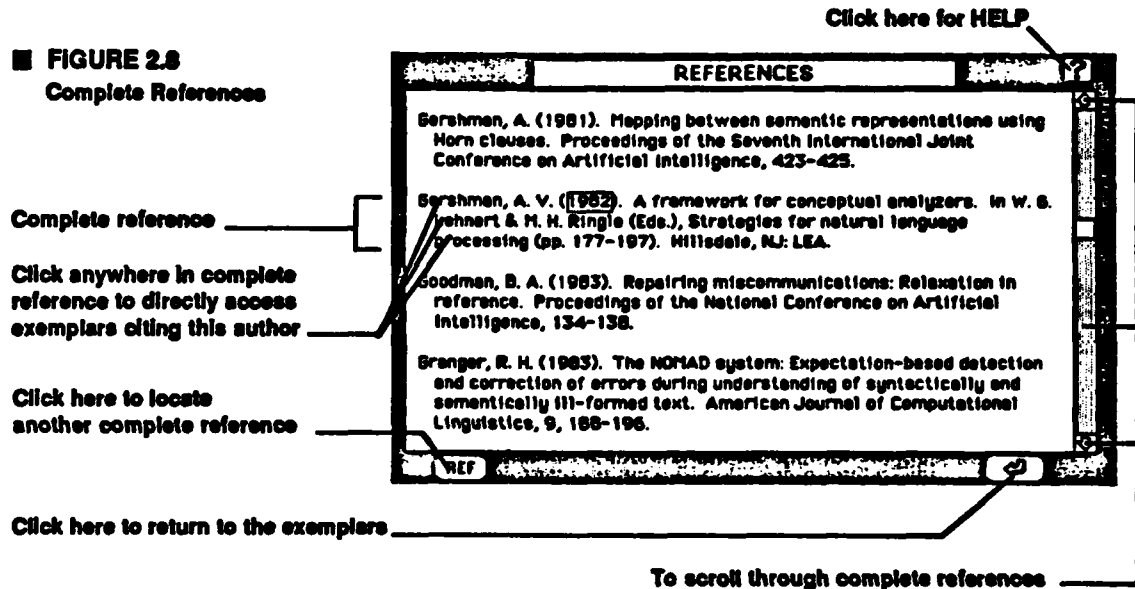
The cognitive-psychological classification scheme is quite lengthy (11 pages). One card was assigned to each of its six major sections (see Figure 2.7 for an example of one of the sections).



- **Complete References**

The complete references are displayed on one card and are contained in one scrolling field (see Figure 2.8).

■ **FIGURE 2.8**
Complete References



Sourcebook Database Browsing and Searching Features

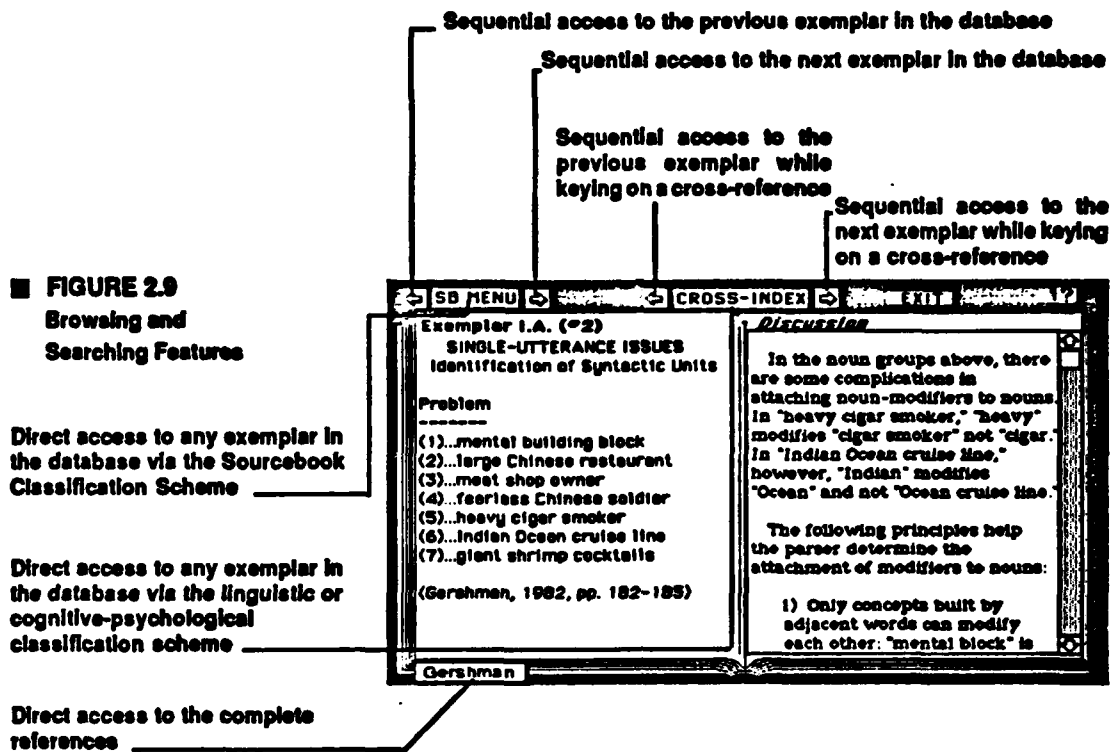
The components of the Natural Language Sourcebook were integrated into the HyperCard environment with the following design enhancements tailored specifically for the Sourcebook Database:

- **Sequential access to the exemplars in the database in the order they are presented in the Natural Language Sourcebook**

The database can be read like the Natural Language Sourcebook. The user can click on either of the browsing buttons at the top left of any exemplar card next to the button labeled "SB MENU" (see Figure 2.9). The user will go to the previous or the next exemplar based upon which button he/she clicked.

- **Direct access to any exemplar in the database based upon its Sourcebook classification scheme**

The user can directly access any exemplar by clicking on the button labeled "SB MENU" at the top left of any exemplar card (see Figure 2.9). The user will access the Sourcebook Classification Menu (see Figure 2.5), and this stepwise search will conclude with the selection of the desired exemplar identification number (see Figure 2.10).



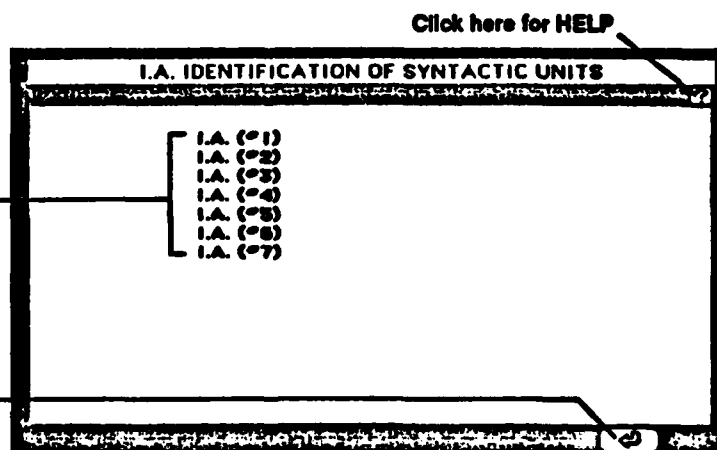
- Direct access to the first exemplar in the database containing a specific linguistic cross-reference or cognitive-psychological cross-reference

The user can access the first exemplar in the database containing a specific linguistic cross-reference or cognitive-psychological cross-reference by clicking on the button labeled "CROSS-INDEX" located at the middle of the top of any typical exemplar (see Figure 2.9). The user will go to the first exemplar in the database containing the classification entry that he/she clicks on while looking at either of the classification schemes (see Figure 2.6 or 2.7).

FIGURE 2.10
Sourcebook
Classification Submenu

Exemplars in I.A.
classification entry

Click here to return
to the exemplars



- Sequential access to exemplars in the database while keying on a specific linguistic classification cross-reference or a cognitive-psychological cross-reference

The user can key through the database on a specific linguistic cross-reference or cognitive-psychological cross-reference by clicking on the left or right cross-index arrow next to the button labeled "CROSS-INDEX" located at the middle of the top of any typical exemplar (see Figure 2.9). The cross-references will be displayed (see Figure 2.4). The user can click on any displayed cross-reference to go to the previous or the next exemplar (depending on which direction arrow was clicked initially).

- Direct access to the complete reference of the current exemplar

The user can access the complete reference of the current exemplar he/she is looking at by clicking on the author's name at the bottom left corner of any exemplar card (see Figure 2.9). The complete reference will appear on the screen along with adjacent references (see Figure 2.8).

- Search procedure for any author in the complete references

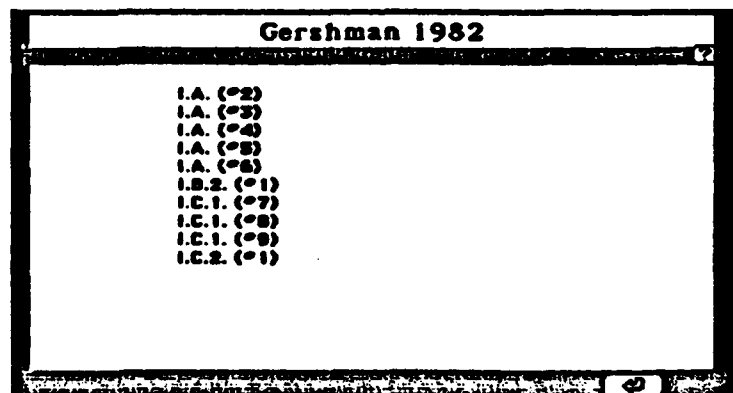
After the user accesses the complete references he/she can access any author in the complete references by click on the button labeled "REF" at the bottom left of the complete references card (see Figure 2.8).

- Direct access to any exemplar citing a specific author

While looking at the complete references, the user can directly access any exemplar in the database that cites a particular reference by clicking anywhere in that reference (see Figure 2.8). A list of exemplars will be displayed, any one of which may be accessed by clicking on the line containing the desired exemplar identification number (see Figure 2.11).

■ FIGURE 2.11
List of Gershman (1982)
Exemplars

Exemplars citing
Gershman (1982)



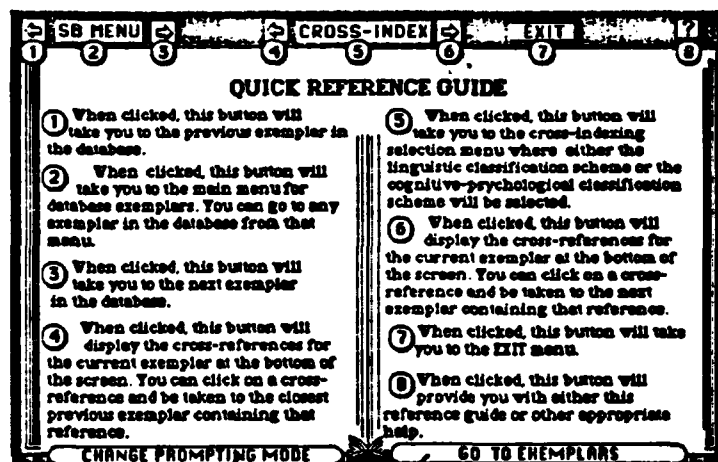
On-Line Assistance

Two types of on-line assistance are available to the user:

- Quick Reference Guide

The Quick Reference Guide describes in detail the functions of each of the buttons that appear at the top of every exemplar screen (see Figure 2.12).

FIGURE 2.12
Quick Reference Guide



Click here to change prompting mode

Click here to go to database exemplars

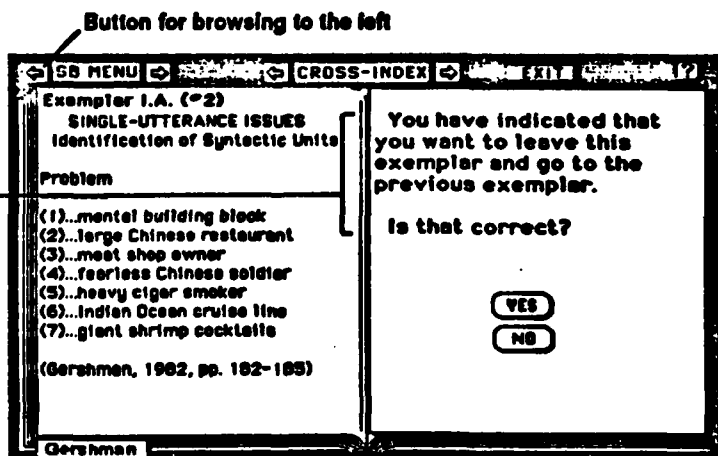
- Prompting Mode

The prompting mode provides explanation of any button clicked by the user. The three types of prompting messages are: 1) YES/NO, 2) Classification, and 3) Cross-Reference. An example of one type of prompting message is shown in Figure 2.13.

FIGURE 2.13
Browsing With Prompting

YES / NO

Prompting message



Sourcebook Database User's Manual

The Sourcebook Database User's Manual provides a step-by-step description of how to access the information in the database version of the Natural Language Sourcebook. The user's manual consists of:

- **Table of Contents**

- **Initial Section**

The initial section provides information about the hardware required to use the database, as well as advice for backing up the database. It provides the description of the the Sourcebook Database User's Manual that is found below in the chapter information section.

- **Four Chapters**

- Chapter 1, "Introduction to the Sourcebook," provides a description and explanation of the Sourcebook.
- Chapter 2, "Browsing Through the Database," describes how to begin using the database. A description of a typical exemplar is provided with explanations of the types of fields (Problem, Discussion, Cross-references) and buttons used to browse through the database exemplars. The method of accessing the references is discussed, and finally, the process of exiting the database is described.
- Chapter 3, "Searching for Exemplars and Cross-References," provides information on searching through the exemplars via the Sourcebook classification scheme, linguistic and cognitive-psychological cross-references and the complete references.
- Chapter 4, "On-Line Assistance," provides a description of the Quick Reference Guide and explains how to use the prompting mode.

- **Appendix**

The appendix provides a discussion of the linguistic and cognitive-psychological classification schemes.

Summary

This report describes the development of a HyperCard stack version of the data presented in the Natural Language Sourcebook. The HyperCard stack version of the Sourcebook capitalizes on the modular structure of the Natural Language Sourcebook exemplars. HyperCard organizational features are appropriate to a functional presentation of the Sourcebook structure. For example, the linking of exemplars by various classification schemes is concretely and visually realized by links between specific cards. The hierarchical nature of HyperCard provides an efficient means of organizing the structure of the exemplars.

The HyperCard Sourcebook database will be an asset to artificial intelligence researchers, cognitive scientists, and computational linguists who wish to access the information contained in the Natural Language Sourcebook.

References

Herl, H. (August 1990). User's Manual, HyperCard Database for the Natural Language Sourcebook. Center for Technology Assessment, UCLA Center for the Study of Evaluation.

Read, W., Dyer, M., Baker, E., Mutch, P., Butler, F., Quilici, A., & Reeves, J. (1990). Natural Sourcebook. Center for Technology Assessment, Center for the Study of Evaluation and Computer Science Department, UCLA.